

0590
1113

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/724,797

DATE: 11/14/2001

TIME: 15:08:34

Input Set : A:\2653-40 Sequence Listing.txt

Output Set: N:\CRF3\11142001\I724797.raw

4 <110> APPLICANT: Jon S. THORSON
6 <120> TITLE OF INVENTION: MICROMONOSPORA ECHINOSPORA GENES
7 ENCODING FOR BIOSYNTHESIS OF
8 CALICHEAMICIN AND SELF-RESISTANCE THERETO
10 <130> FILE REFERENCE: 2653-40
12 <140> CURRENT APPLICATION NUMBER: 09/724,797
13 <141> CURRENT FILING DATE: 2000-11-28
15 <150> PRIOR APPLICATION NUMBER: 60/111,325
16 <151> PRIOR FILING DATE: 1998-12-07
18 <160> NUMBER OF SEQ ID NOS: 95
20 <170> SOFTWARE: FastSEQ for Windows Version 4.0
22 <210> SEQ ID NO: 1
23 <211> LENGTH: 546
24 <212> TYPE: DNA
25 <213> ORGANISM: Bacteria
27 <220> FEATURE:
28 <221> NAME/KEY: CDS
29 <222> LOCATION: (1)...(546)
31 <400> SEQUENCE: 1
32 atg act cag gag aag acc gca ccg gcc gcg aag agc acg acc acc aag 48
33 Met Thr Gln Glu Lys Thr Ala Pro Ala Ala Lys Ser Thr Thr Thr Lys
34 1 5 10 15
36 agc acc gcc gcg aag aag ccg aag ccc ccg aac tac gac ccg ttc gtc 96
37 Ser Thr Ala Ala Lys Lys Pro Lys Pro Pro Asn Tyr Asp Pro Phe Val
38 20 25 30
40 cgg cac agc gtc act gtc aag gcc gac cgc aag acc gcc ttc aag acg 144
41 Arg His Ser Val Thr Val Lys Ala Asp Arg Lys Thr Ala Phe Lys Thr
42 35 40 45
44 ttc ctc gaa ggc ttt ccg gag tgg tgg ccg aac aac ttc cgc acc acc 192
45 Phe Leu Glu Gly Phe Pro Glu Trp Trp Pro Asn Asn Phe Arg Thr Thr
46 50 55 60
48 aag gtc ggg gcc ccg ctg ggc gtc gac aag aag ggc ggc cgc tgg tac 240
49 Lys Val Gly Ala Pro Leu Gly Val Asp Lys Lys Gly Gly Arg Trp Tyr
50 65 70 75 80
52 gag atc gac gag cag ggc gag gag cac acc ttc ggc ctg atc cgg aag 288
53 Glu Ile Asp Glu Gln Gly Glu Glu His Thr Phe Gly Leu Ile Arg Lys
54 85 90 95
56 gtg gac gag ccg gac acg ctg gtc atc ggc tgg cgg ctc aac ggc ttc 336
57 Val Asp Glu Pro Asp Thr Leu Val Ile Gly Trp Arg Leu Asn Gly Phe
58 100 105 110
60 ggc cgg atc gac ccg gac aac tcg agc gag ttc acc gtg acc ttc gtg 384
61 Gly Arg Ile Asp Pro Asp Asn Ser Ser Glu Phe Thr Val Thr Phe Val
62 115 120 125
64 gcc gac ggc cag aag aag acc cgg gtg gac gtc gag cac acc cac ttc 432
65 Ala Asp Gly Gln Lys Lys Thr Arg Val Asp Val Glu His Thr His Phe
66 130 135 140
68 gac cgg atg ggc acc aag cac gcc aag cgg gtc cgc aac ggc atg gac 480

ENTERED

P.5

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/724,797

DATE: 11/14/2001

TIME: 15:08:34

Input Set : A:\2653-40 Sequence Listing.txt

Output Set: N:\CRF3\11142001\I724797.raw

```

69 Asp Arg Met Gly Thr Lys His Ala Lys Arg Val Arg Asn Gly Met Asp
70 145          150          155          160
72 aag ggc tgg ccg acg atc ctc cag tcg ttc cag gac aag atc gac gag 528
73 Lys Gly Trp Pro Thr Ile Leu Gln Ser Phe Gln Asp Lys Ile Asp Glu
74          165          170          175
76 gaa ggg gcg aag aag tga 546
77 Glu Gly Ala Lys Lys *
78          180
81 <210> SEQ ID NO: 2
82 <211> LENGTH: 181
83 <212> TYPE: PRT
84 <213> ORGANISM: Bacteria
86 <400> SEQUENCE: 2
87 Met Thr Gln Glu Lys Thr Ala Pro Ala Ala Lys Ser Thr Thr Thr Lys
88 1          5          10          15
89 Ser Thr Ala Ala Lys Lys Pro Lys Pro Pro Asn Tyr Asp Pro Phe Val
90          20          25          30
91 Arg His Ser Val Thr Val Lys Ala Asp Arg Lys Thr Ala Phe Lys Thr
92          35          40          45
93 Phe Leu Glu Gly Phe Pro Glu Trp Trp Pro Asn Asn Phe Arg Thr Thr
94          50          55          60
95 Lys Val Gly Ala Pro Leu Gly Val Asp Lys Lys Gly Gly Arg Trp Tyr
96 65          70          75          80
97 Glu Ile Asp Glu Gln Gly Glu Glu His Thr Phe Gly Leu Ile Arg Lys
98          85          90          95
99 Val Asp Glu Pro Asp Thr Leu Val Ile Gly Trp Arg Leu Asn Gly Phe
100          100          105          110
101 Gly Arg Ile Asp Pro Asp Asn Ser Ser Glu Phe Thr Val Thr Phe Val
102          115          120          125
103 Ala Asp Gly Gln Lys Lys Thr Arg Val Asp Val Glu His Thr His Phe
104          130          135          140
105 Asp Arg Met Gly Thr Lys His Ala Lys Arg Val Arg Asn Gly Met Asp
106 145          150          155          160
107 Lys Gly Trp Pro Thr Ile Leu Gln Ser Phe Gln Asp Lys Ile Asp Glu
108          165          170          175
109 Glu Gly Ala Lys Lys
110          180
113 <210> SEQ ID NO: 3
114 <211> LENGTH: 1155
115 <212> TYPE: DNA
116 <213> ORGANISM: Bacteria
118 <220> FEATURE:
119 <221> NAME/KEY: CDS
120 <222> LOCATION: (1)...(1155)
122 <400> SEQUENCE: 3
123 atg gca act agc gag agg ggt gtc atg atc ccg ctg tcc aag gtc gcc 48
124 Met Ala Thr Ser Glu Arg Gly Val Met Ile Pro Leu Ser Lys Val Ala
125 1          5          10          15
127 atg tct ccg gac gtc agc acc cgc gtc tcc gcc gtc ctg agc agt ggc 96

```

RAW SEQUENCE LISTING

DATE: 11/14/2001

PATENT APPLICATION: US/09/724,797

TIME: 15:08:34

Input Set : A:\2653-40 Sequence Listing.txt

Output Set: N:\CRF3\11142001\I724797.raw

```

128 Met Ser Pro Asp Val Ser Thr Arg Val Ser Ala Val Leu Ser Ser Gly
129                20                25                30
131 cgg ctg gag cac ggg ccg acc gtc gcc gag tac gag gcg gcc gtg ggc 144
132 Arg Leu Glu His Gly Pro Thr Val Ala Glu Tyr Glu Ala Ala Val Gly
133                35                40                45
135 agt cgt atc ggc aac ccc cgg gtg gtc tcg gtc aac tgc ggc acg gcc 192
136 Ser Arg Ile Gly Asn Pro Arg Val Val Ser Val Asn Cys Gly Thr Ala
137                50                55                60
139 ggg ctc cac ctg gcg ctg agc ctc gcc gcg cgg ccg ggg gcc ggc gag 240
140 Gly Leu His Leu Ala Leu Ser Leu Ala Ala Arg Pro Gly Ala Gly Glu
141        65                70                75                80
143 tcg gag cac gac ggc ccg ggc gag gtg ctc acc acg ccg ctg acc ttc 288
144 Ser Glu His Asp Gly Pro Gly Glu Val Leu Thr Thr Pro Leu Thr Phe
145                85                90                95
147 gag ggc acg aac tgg ccg atc ctc gcc aac ggg ctg cgc atc cgg tgg 336
148 Glu Gly Thr Asn Trp Pro Ile Leu Ala Asn Gly Leu Arg Ile Arg Trp
149                100               105               110
151 gtg gac gtc gac ccg gcc acc ctc aac atg gac ctc gac gac ctg gcc 384
152 Val Asp Val Asp Pro Ala Thr Leu Asn Met Asp Leu Asp Asp Leu Ala
153                115               120               125
155 gcg aag atc tcg ccc gcc acc cgg gcc atc gtg gtg gtc cac tgg ctc 432
156 Ala Lys Ile Ser Pro Ala Thr Arg Ala Ile Val Val Val His Trp Leu
157        130               135               140
159 ggc tac ccg gtg gac ctc aac cgg ctg cgc gcc gtc gtg gac cgg gcc 480
160 Gly Tyr Pro Val Asp Leu Asn Arg Leu Arg Ala Val Val Asp Arg Ala
161 145                150                155                160
163 acg gcg gga tac gac cgc cgc ccg ctg gtc gtg gag gac tgc gcg cag 528
164 Thr Ala Gly Tyr Asp Arg Arg Pro Leu Val Val Glu Asp Cys Ala Gln
165                165               170               175
167 gcg tgg ggc gcc acc tac cgg ggc gcg ccg ctg ggc acg cac ggc aac 576
168 Ala Trp Gly Ala Thr Tyr Arg Gly Ala Pro Leu Gly Thr His Gly Asn
169                180               185               190
171 gtc tgc gtg tac agc acc ggc gcg atc aag atc ctg acg acc ggc agc 624
172 Val Cys Val Tyr Ser Thr Gly Ala Ile Lys Ile Leu Thr Thr Gly Ser
173                195               200               205
175 ggc ggc ttc gtc gtg ctg ccc gac gac gac ctg tac gac cgg ctc cgg 672
176 Gly Gly Phe Val Val Leu Pro Asp Asp Asp Leu Tyr Asp Arg Leu Arg
177        210               215               220
179 ctg cgc cgc tgg ctc ggc atc gag cgg gcg tcg gac cgg atc acc ggc 720
180 Leu Arg Arg Trp Leu Gly Ile Glu Arg Ala Ser Asp Arg Ile Thr Gly
181 225                230                235                240
183 gac tac gac gtc gcc gag tgg ggc tac cgg ttc atc ctc aac gag atc 768
184 Asp Tyr Asp Val Ala Glu Trp Gly Tyr Arg Phe Ile Leu Asn Glu Ile
185                245                250                255
187 ggc ggg gcg atc ggc ctg tcc aac ctg gaa cgc gtc gac gag ctg ctg 816
188 Gly Gly Ala Ile Gly Leu Ser Asn Leu Glu Arg Val Asp Glu Leu Leu
189                260                265                270
191 cgc cgg cac cgg gag aac gcc gcg ttc tac gac aag gaa ctg gcc ggc 864
192 Arg Arg His Arg Glu Asn Ala Ala Phe Tyr Asp Lys Glu Leu Ala Gly

```

RAW SEQUENCE LISTING

DATE: 11/14/2001

PATENT APPLICATION: US/09/724,797

TIME: 15:08:34

Input Set : A:\2653-40 Sequence Listing.txt

Output Set: N:\CRF3\11142001\I724797.raw

```

193          275          280          285
195 atc gac ggc gtc gag cag acc gag cgg gcc gac gac cgg gag ccc gcg 912
196 Ile Asp Gly Val Glu Gln Thr Glu Arg Ala Asp Asp Arg Glu Pro Ala
197          290          295          300
199 ttc tgg atg tac ccg ctg aag gtc cgc gac cgt ccc gcc ttc atg cgc 960
200 Phe Trp Met Tyr Pro Leu Lys Val Arg Asp Arg Pro Ala Phe Met Arg
201 305          310          315          320
203 cgg ctg ctc gac gcc ggc atc gcc acc agc gtc gtg tcg cgc cgc aac 1008
204 Arg Leu Leu Asp Ala Gly Ile Ala Thr Ser Val Val Ser Arg Arg Asn
205          325          330          335
207 gac gcg cac agc tgc gtc gcg tcg gcc cgc acc acc ctg ccc ggg ctg 1056
208 Asp Ala His Ser Cys Val Ala Ser Ala Arg Thr Thr Leu Pro Gly Leu
209          340          345          350
211 gac cgg gtg gcg gac cgc gtg gtc cac atc ccg gtg ggc tgg tgg ctc 1104
212 Asp Arg Val Ala Asp Arg Val Val His Ile Pro Val Gly Trp Trp Leu
213          355          360          365
215 acc gag gac gac cgc tcc cac gtc gtc gaa acg atc aag tcc ggc tgg 1152
216 Thr Glu Asp Asp Arg Ser His Val Val Glu Thr Ile Lys Ser Gly Trp
217          370          375          380
219 tga 1155
220 *
224 <210> SEQ ID NO: 4
225 <211> LENGTH: 384
226 <212> TYPE: PRT
227 <213> ORGANISM: Bacteria
229 <400> SEQUENCE: 4
230 Met Ala Thr Ser Glu Arg Gly Val Met Ile Pro Leu Ser Lys Val Ala
231 1 5 10 15
232 Met Ser Pro Asp Val Ser Thr Arg Val Ser Ala Val Leu Ser Ser Gly
233 20 25 30
234 Arg Leu Glu His Gly Pro Thr Val Ala Glu Tyr Glu Ala Ala Val Gly
235 35 40 45
236 Ser Arg Ile Gly Asn Pro Arg Val Val Ser Val Asn Cys Gly Thr Ala
237 50 55 60
238 Gly Leu His Leu Ala Leu Ser Leu Ala Ala Arg Pro Gly Ala Gly Glu
239 65 70 75 80
240 Ser Glu His Asp Gly Pro Gly Glu Val Leu Thr Thr Pro Leu Thr Phe
241 85 90 95
242 Glu Gly Thr Asn Trp Pro Ile Leu Ala Asn Gly Leu Arg Ile Arg Trp
243 100 105 110
244 Val Asp Val Asp Pro Ala Thr Leu Asn Met Asp Leu Asp Asp Leu Ala
245 115 120 125
246 Ala Lys Ile Ser Pro Ala Thr Arg Ala Ile Val Val Val His Trp Leu
247 130 135 140
248 Gly Tyr Pro Val Asp Leu Asn Arg Leu Arg Ala Val Val Asp Arg Ala
249 145 150 155 160
250 Thr Ala Gly Tyr Asp Arg Arg Pro Leu Val Val Glu Asp Cys Ala Gln
251 165 170 175
252 Ala Trp Gly Ala Thr Tyr Arg Gly Ala Pro Leu Gly Thr His Gly Asn

```

RAW SEQUENCE LISTING

DATE: 11/14/2001

PATENT APPLICATION: US/09/724,797

TIME: 15:08:34

Input Set : A:\2653-40 Sequence Listing.txt

Output Set: N:\CRF3\11142001\I724797.raw

```

253          180          185          190
254 Val Cys Val Tyr Ser Thr Gly Ala Ile Lys Ile Leu Thr Thr Gly Ser
255          195          200          205
256 Gly Gly Phe Val Val Leu Pro Asp Asp Asp Leu Tyr Asp Arg Leu Arg
257          210          215          220
258 Leu Arg Arg Trp Leu Gly Ile Glu Arg Ala Ser Asp Arg Ile Thr Gly
259 225          230          235          240
260 Asp Tyr Asp Val Ala Glu Trp Gly Tyr Arg Phe Ile Leu Asn Glu Ile
261          245          250          255
262 Gly Gly Ala Ile Gly Leu Ser Asn Leu Glu Arg Val Asp Glu Leu Leu
263          260          265          270
264 Arg Arg His Arg Glu Asn Ala Ala Phe Tyr Asp Lys Glu Leu Ala Gly
265          275          280          285
266 Ile Asp Gly Val Glu Gln Thr Glu Arg Ala Asp Asp Arg Glu Pro Ala
267          290          295          300
268 Phe Trp Met Tyr Pro Leu Lys Val Arg Asp Arg Pro Ala Phe Met Arg
269 305          310          315          320
270 Arg Leu Leu Asp Ala Gly Ile Ala Thr Ser Val Val Ser Arg Arg Asn
271          325          330          335
272 Asp Ala His Ser Cys Val Ala Ser Ala Arg Thr Thr Leu Pro Gly Leu
273          340          345          350
274 Asp Arg Val Ala Asp Arg Val Val His Ile Pro Val Gly Trp Trp Leu
275          355          360          365
276 Thr Glu Asp Asp Arg Ser His Val Val Glu Thr Ile Lys Ser Gly Trp
277          370          375          380
280 <210> SEQ ID NO: 5
281 <211> LENGTH: 990
282 <212> TYPE: DNA
283 <213> ORGANISM: Bacteria
285 <220> FEATURE:
286 <221> NAME/KEY: CDS
287 <222> LOCATION: (1)...(990)
288 <223> OTHER INFORMATION: biosynthetic gene
290 <400> SEQUENCE: 5
291 gtg ccc aga tcc ctg gtc acc ggc ggc ttc ggc ttc gtc ggc agt cac 48
292 Val Pro Arg Ser Leu Val Thr Gly Gly Phe Gly Phe Val Gly Ser His
293 1 5 10 15
295 gtc gtc gaa cgg ctg gtc cgc cgg ggt gac gag gtc gtc gtc tac gac 96
296 Val Val Glu Arg Leu Val Arg Arg Gly Asp Glu Val Val Val Tyr Asp
297 20 25 30
299 ctc gcc gac ccg ccg ccc gac ctg gag cac ccg ccg ggc gcg atc cgg 144
300 Leu Ala Asp Pro Pro Pro Asp Leu Glu His Pro Pro Gly Ala Ile Arg
301 35 40 45
303 cac gtc cgc ggc gac gtc cgg gac gcc gac ggg ctg gcg gcc gcc gcc 192
304 His Val Arg Gly Asp Val Arg Asp Ala Asp Gly Leu Ala Ala Ala Ala
305 50 55 60
307 acc ggc gtg gac gag gtc tac cac ctc gcg gcg gtc gtc ggc gtc gac 240
308 Thr Gly Val Asp Glu Val Tyr His Leu Ala Ala Val Val Gly Val Asp
309 65 70 75 80

```

Use of n and / or Xaa has been detected in the Sequence Listing. Review the Sequence Listing to ensure a corresponding explanation is present in the <220> to <223> fields of each sequence using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/724,797

DATE: 11/14/2001

TIME: 15:08:35

Input Set : A:\2653-40 Sequence Listing.txt

Output Set: N:\CRF3\11142001\I724797.raw

L:649 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9
L:650 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9
L:779 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10